Data-driven Background Modeling for the Self-Coupling of the Higgs Boson

Setayesh Nekarae[1], John Alison[2], Mikael Kuusela[3]

Department of Physics and Astronomy, [1] University of California, Los Angeles, [2] Carnegie Mellon University
[3] Department of Statistics, Carnegie Mellon University

Introduction

The Large Hadron Collider (LHC) at CERN is a 27km particle accelerator with superconducting magnets that collide protons at **99.9% speed of the light** creating jets [1]. Of particular interest is analyzing data to detect Di-Higgs events as their existence can confirm the theory of **mass generation of the Higgs boson** [2].

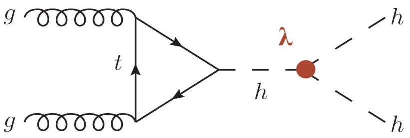

Figure 1. Process of Di-Higgs events creation at CERN. Image reproduced from [3].

There is **1 Higgs boson produced per billion collider events** which cannot be detected directly due to its instability. Thus our aim is to detect Di-Higgs production from its most probabilistic decay to b-quarks [3].

Analysis Methods

A collider event can be represented by a discrete measure,

$$g = \sum_{i=1}^{K} p_{T_i} \delta_{(\eta_i, \phi_i, m_i)}$$

Ref. [4]

Higgs bosons **attain their mass from the opening angle** of the corresponding b-quark decay.

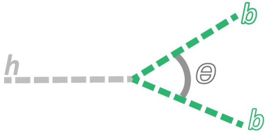

Figure 2. Higgs decay to b-quarks from which it attains its mass.

Namely this is governed by,

$$m^2_{\text{Higgs}} = 2|\vec{p_{b,1}}||\vec{p_{b,2}}|(1 - \cos\theta)$$

$$\theta = \sqrt{\Delta\phi^2 + \Delta\eta^2}$$

Results

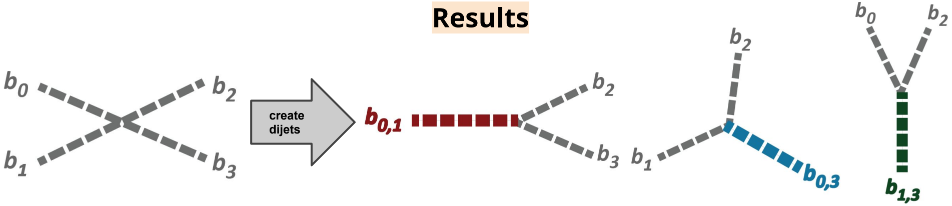

Figure 3. Possible dijet pairings from which the mass can be deduced.

Dijet masses for all possible pairings were calculated and plotted with respect to the addition of their scalar momentas. ''Signal-like' events were filtered out based on their distance to the known Higgs mass. Histograms were computed comparing signal and background-like events to verified signal events to estimate efficacy of this method.

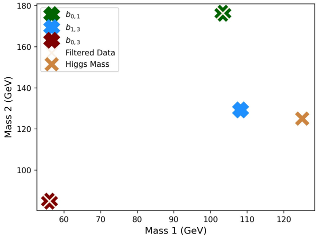

Figure 4. Display of possible dijet masses for a single event. The filtered data represents dataset that can be tagged as 'background' while the point closest to the Higgs mass is tagged as a potential signal. This process was repeated for the entire dataset (on the order of hundreds of thousand of events.

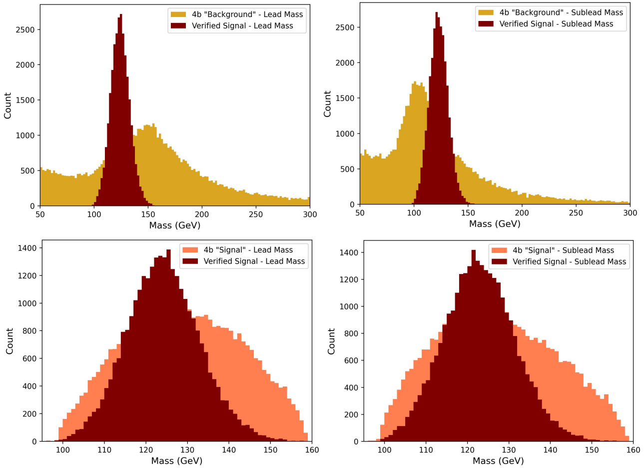

Figure 5. Histograms depicting signal and background-like events compared to verified signal events.

Future endeavors

Future plans include training a classifier that would be able to more definitively isolate signal events by estimating the background. Additionally, we would like to account for associated uncertainties in the measurements.

References

[1] The Large Hadron Collider. CERN.
[2] Martin, B. R., & Shaw, G. Particle Physics (Section 10.2 Gauge invariance and the Higgs boson). Wiley.
[3] Alison, J. The Discovery of the Higgs Boson.
[4] Wasserman, L. et al. Optimal Transport With Applications to Background Modeling. CERN.

Acknowledgments

I would like to thank Prof. Alison and Prof. Kuusela for their guidance and mentorship throughout the duration of this program. Thank you to the CMU SURP NSF AI Planning Institute for Data-Driven Discovery for funding this work.